Author’s response to reviews

Title: A Clinical Prediction Rule for Diagnosing Human Infections with Avian Influenza A(H7N9) in a Hospital Emergency Department Setting

Authors:

Qiaohong Liao (liaoqh@chinacdc.cn)
Dennis K. M. Ip (dkmip@hku.hk)
Tim K. Tsang (matklab@gmail.com)
Bin Cao (caobin1999@gmail.com)
Hui Jiang (jianghui@chinacdc.cn)
Fengfeng Liu (liuff@chinacdc.cn)
Jiandong Zheng (zhengjd@chinacdc.cn)
Zhibin Peng (pengzb@chinacdc.cn)
Peng Wu (pengwu@hku.hk)
Yang Huai (huaiy@CN.CDC.GOV)
Eric H. Y. Lau (ehylau@hku.hk)
Luzhao Feng (fenglz@chinacdc.cn)
Gabriel M. Leung (gmleung@hku.hk)
Hongjie Yu (yuhi@chinacdc.cn)
Benjamin J. Cowling (bcowling@hku.hk)

Version: 3
Date: 10 May 2014

Author’s response to reviews:

May 02, 2014

John Bartlett
Editor-in-Chief
BMC Medicine

Dear Prof. Bartlett,

RE: A Clinical Prediction Rule for Diagnosing Human Infections with Avian Influenza A(H7N9) in a Hospital Emergency Department Setting

Thank you for considering our manuscript for publication in BMC Medicine. Your comments and those from the reviewers have helped us to revise and improve our manuscript. I have enclosed the revised manuscript and below I have detailed the point-by-point responses to the reviewers’ comments.

In response to the comments by both reviewers’ we have added one additional figure (Figure 2) to better elaborate on the performance of our prediction rule, so as to convince the reviewers and the readers that we had chosen the model with the best performance in terms of both the high sensitivity and area under the
ROC curve. We also added an Appendix 1 to give more details on the multiple imputation and validation processes involved in building the risk prediction model.

On behalf of all authors,
Yours sincerely,
Hongjie Yu, MD PhD MPH
Division of Infectious Disease
Key Laboratory of Surveillance and Early-warning on Infectious Disease, China CDC
155# Changbai Road, Beijing, 102206, P.R. China
yuhj@chinacdc.cn; +86 10 58900548 (Tel); +86 10 58900548 (fax)

Point-by-point response to reviewers' comment

Reviewer 1: Stefano Merler
Reviewer's report:
In this paper authors describe a method to timely discriminate between A/H7N9 cases and ARI based on clinical variables readily available soon after hospital admission. They claim that the proposed method could help physicians to optimize resources, especially in case of a new epidemic wave. The paper is interesting (in particular I like the idea of simplifying results by transforming model coefficient into scores, easy to use by physicians), the performed statistical analysis sounds correct to me, but there a few points that should be further discussed/clarified.

Major Compulsory Revisions:
The major problem I have with this paper is that authors do not demonstrate that performances (sensitivity of 0.93, a specificity of 0.80, or 0.96-0.75 as resulting from bootstrapping) are somehow optimal and may be really helpful to optimize resources during an epidemic. Which is the overall cost of misdiagnosing 7% of A/H7N9 cases? Which is the cost of misclassifying 20% of non A/H7N9 ARI cases? In general the answer to the last question strongly depends on the overall incidence of ARI cases. In fact, which is the gain if the number of ARI is so large that 20% of them make the proposed prediction rule of unpractical use? In principle, one might be happier with sensitivity of 1 at the price, of course, of specificity lower than 0.8, or specificity 1, with sensitivity lower than 0.93.

Response 1: As A/H7N9 infection is being generally more severe clinically and associated with more complications, its early identification for early treatment to prevent progression is of utmost importance, and the cost for missing an A/H7N9 case is probably much larger than misclassifying a non A/H7N9 case. The decision rule is mainly aim to be used as a screening tool to capture most of the A/H7N9 cases if possible. As most other screening tools, we therefore selected the model with a higher (95%) sensitivity (93% after correction), with the possible tradeoff of a less then perfect specificity. We agreed that a 20% false positive
rate may still carry a non-trivial resource implication when the overall ARI incidence is high. However, there should still be a sizable gain when compared with a scenario where the rule is lacking and resources are quickly misdirected for the management of an even larger proportion of misclassified cases. And this is exactly the reason why the decision rule should be a helpful tool for us in the frontline.

Related to the previous point, I think that showing the entire sensitivity-specificity curve, and discussing different regions of the curve, could be more informative. Afterwards, you could still claim that the model with 0.93-0.8 sen-spe is the optimal one. Moreover, there do exist techniques, based on weighting (basically you may give more importance to A/H7N9 cases or to ARI cases), that allow us to get higher specificity or sensitivity.

Response 2: We agree this curve would be useful and we added the sensitivity-specificity curve (ROC curve) as Figure 2 on page 14, paragraph 1, line 2. As we considered missing an A/H7N9 case is much serious compared with misclassifying a non-A/H7N9 ARI case, the cutoff of the model is chosen to have at least 95% (with bootstrap correction 93%) sensitivity so as to correctly and timely identify as many A/H7N9 cases as possible. From the ROC curve, it can be shown that the optimal combination of performance has been chosen in our model, comparing to either the scenario where, the specificity is only about 60% when we have perfect sensitivity (100%), or the sensitivity is only about 70% when we have perfect specificity (100%).

Minor Essential Revisions:

Authors claim that the proposed method could be helpful in the case of the re-emergence of the epidemic. I think they should clarify that they are still thinking of a virus with little or no human-to-human transmission potential. Should A/H7N9 evolve into a (readily) human-to-human transmissible virus, the proposed prediction rule is of little value (the most important variable is “history of poultry exposure”).

Response 3: We agreed with this important comment and have this clarified in the limitation on page 17, paragraph 1, lines 16-19.

It is not clear to me why authors excluded a priori from the study individuals aged less than 14 years. They say that only few cases have been observed in this age class (and of course this is true). I think that a possible explanation is that elderly have a higher probability to be exposed to infected live poultry (e.g. in the markets). The point, however, is that variable age could be discarded by the variable selection procedure, even if the 0-14 age group is considered, because “history of poultry exposure” already explains the observed profile of infections by age. This is stronger that excluding a priori the 0-14 age group from the study.

Response 4: Besides a possible difference in the exposure history, important differences of individuals aged less than 14 years from adults and elderly persons may also existed on their presenting symptoms, the approach of symptom ascertainment, overall clinical course and disease severity, disease epidemiology, and even health care seeking pattern and pathway, therefore we
considered they may better be treated as a completely different group and not being included in the current model. We have this further clarified in the revised manuscript on page 7, paragraph 2, lines 7-8.

I think that a brief appendix reporting a more technical description of methods, e.g. multiple imputation, optimism-corrected estimates, etc. would be helpful.

Response 5: We agree with this comment and added a brief appendix as advised on page 33.

Discretionary Revisions:
As for the predictive ability of the proposed method, perhaps authors could be able to estimate performances on new A/H7N9 cases (Jan.-Feb. 2014).

Response 6: This is a very reasonable suggestion but sadly clinical data in the level of detail that is needed in the model from new A/H7N9 cases is not currently available to us.

Reviewer 2: Marianne A.B. van der Sande

Reviewer's report:
Major comments:
1. The authors assume that clinical presentation of those with A/H7 in this first wave will be identical to the clinical presentation in subsequent waves. Previous influenza outbreaks have shown that severity can vary significantly between waves (for better and for worse). Also, health care seeking behaviour (and thus who presents for evaluation or not) can vary significantly over time, also related to social, non-medical reasons. The authors should reflect on this limitation in time and place of their cases as it might also limit the applicability of their prediction tool in another time and place, eg the evolving epidemic the authors refer to.

Response 7: We agreed with this comment, which may be a limitation in the applicability of our model, and have this highlighted in the limitations on page 17, paragraph 1, lines 19-21.

2. I appreciate the authors used the historical data available as controls, but could another option have been to apply (also) a “test-negative” approach (as is frequently used to estimate influenza vaccine effectiveness) and compare clinical details of those tested pos and those tested neg to assess if this can be used to construct a prediction rule? Unlike with the current selection of controls, cases and controls will then come from the same population. This might be closer to the actual clinical situation this prediction rule is aimed at, where clinicians are faced with a patient suspected to be infected with A/H7 or not, and therefore tested.

Response 8: We thank for this good idea. However, detailed clinical data from test negative cases were not systematically collected like what had been done for positive cases during the period of the epidemic in most hospitals, and thus not readily available at the present stage. We do consider this suggestion a possible further step in further work.
3. Rather than restrict themselves to an internal bootstrap validation, why did the authors not also test their prediction rule on patients admitted during the winter of 2013, when many new cases were reported? Is this still possible?

Response 9: As mentioned in Response 6, this is a very reasonable suggestion but sadly clinical data in the level of detail that is needed in the model from new A/H7N9 cases is not available to us.

4. The authors indicate that they excluded several laboratory variables if 40% or more were missing; table 1 shows several variables included with 33% missing; and the methods mention that multiple imputations were used to make the most of all (?) available non-missing data. Please clarify why this cut off was chosen (and thereby excluding potential information from the available non-missing lab data) and whether or not sensitivity analyses were done to test this?

Response 10: We adopted an arbitrary cut-off of 40% as we feel that >40% may be too much to impute. This also allowed us to include data on exposure history because we thought it should be at least be considered in any reasonable model regarding an infection of a primarily animal origin. In a sensitivity analysis where C-reactive protein (with a missing rate of 50%) had also been considered, it has finally been excluded and didn't result in a great difference in the final model and results.

5. Could the authors present/discuss to what extent selection of different cut-offs and of a different calculation of risk scores might have modified their conclusions.

Response 11: As mentioned in Response 2, we added the ROC curve (sensitivity vs 1-specificity) as Figure 2 to assist the interpretation on the results on page 14, paragraph 1, line 2. Score calculation is based on the result from the regression model and variable selections. As we considered missing an A/H7N9 case is much serious compared with misclassifying a non-A/H7N9 ARI case, the cutoff of the model is chosen to have at least 95% (with bootstrap correction 93%) sensitivity so as to correctly and timely identify as many A/H7N9 cases as possible. From the ROC curve, it can be shown that our chosen is having the optimal combination of performance with a 96% AUC, meaning that comparing with other model lower AUC, our model would give the highest specificity with 95% sensitivity. Adopting a higher cut-off score would improve specificity with less people being falsely classified as A/H7N9 infection with the tradeoff that sensitivity would decrease, and the vice versa for a lower cut-off score.

6. I was not clear how the categorisation of risk scores into tertiles above the cut-off (76) was defined. The risk scores in the tertiles were 76-80; 81-100; >100, the number of patients in each tertile was 48, 362, 223. Please clarify?

Response 12: The number of patients in the three tertiles were 48, 362, 223, this is because 215 patients, being male with fever and radiological evidence of pneumonia/consolidation, were having a score of 71 and just crossing the cut-off between the first two tertiles, thus giving the impression of a large number of people falling into the middle tertile. In the revised manuscript we had also update our tertile cut-offs as 68-70; 71-90; >90 on page 13, paragraph 2, lines
5-7, figure legend of Figure 1, by incorporating newer results of the model basing on a slightly updated data set.

Minor comments:
1. In the conclusions, the authors refer to a potential re-emergence of A/H7 this winter: this has happened and is no longer potential; this can better be included in the background.
   Response 13: Updated as advised on page 18, paragraph 2, line 5 and page 5, paragraph 1, line 7.

2. Typo in heading Table 4: Indies=Indices
   Response 14: Corrected on page 34.